Autonomous Infrastructure for Science: Challenges and Opportunities

Raj Kettimuthu

Argonne National Laboratory and University of Chicago

Team

Tekin Bicer – Assistant Scientist, DSL and APS

Zhengchun Liu – Research Scientist, UChicago/Argonne

Joaquin Chung – Postdoc, DSL

Sam Nickolay – PhD student, UChicago

Pete Beckman – Argonne/Northwestern

Ian Foster – Argonne/UChicago

Autonomous cars (Self driving cars)

Desert race too tough for robots

2005: 5 out of 23 teams completed it

Ranking	Team Name	Hometown		GCE Distance Completed	Average Speed
1	Stanford Racing	Stanford	CA	132 miles	19.1 mph
2	Red Team	Pittsburgh	PA	132	18.6
3	Red Team Too	Pittsburgh	PA	132	18.2
4	Gray Team	Metairie	LA	132	17.5
5	Team TerraMax	Oshkosh	WI	132	10.2
6	Team ENSCO	Springfield	VA	81	
7	Axion Racing	Westlake Village	CA	66	
8	Virginia Tech	Blacksburg	VA	44	
9	Virginia Tech Rocky	Blacksburg	VA	39	
10	Desert Buckeyes	Columbus	OH	29	
11	Insight Racing	Cary	NC	26	
12	Team DAD	Morgan Hill	CA	26	
13	Mojavaton	Grand Junction	CO	23	
14	Golem Group / UCLA	Santa Monica	CA	22	
15	Team CajunBot	Lafayette	LA	17	
16	SciAutonics/Auburn Eng.	Thousand Oaks	CA	16	
17	CIMAR	Gainesville	FL	14	
18	IVST I	Littleton	CO	14	
19	Princeton University	Princeton	NJ	10	
20	Team Cornell	Ithaca	NY	9	
21	Team Caltech	Pasadena	CA	8	
22	MonsterMoto	Cedar Park	TX	7	
23	MITRE Meteorites	McLean	VA	1	

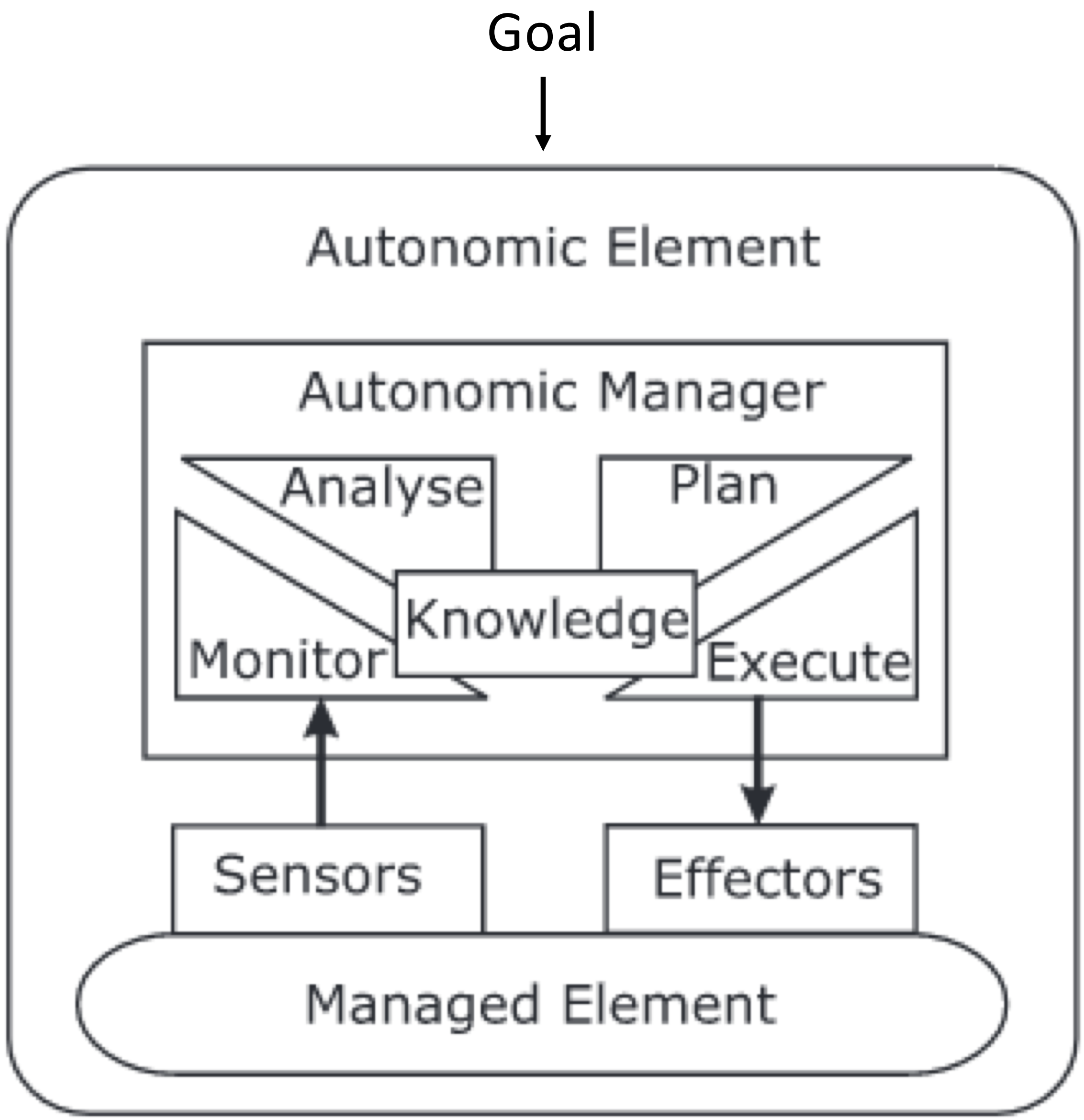
2007: Urban challenge - drive a 96km urban course

Google car - no steering wheel, no pedals

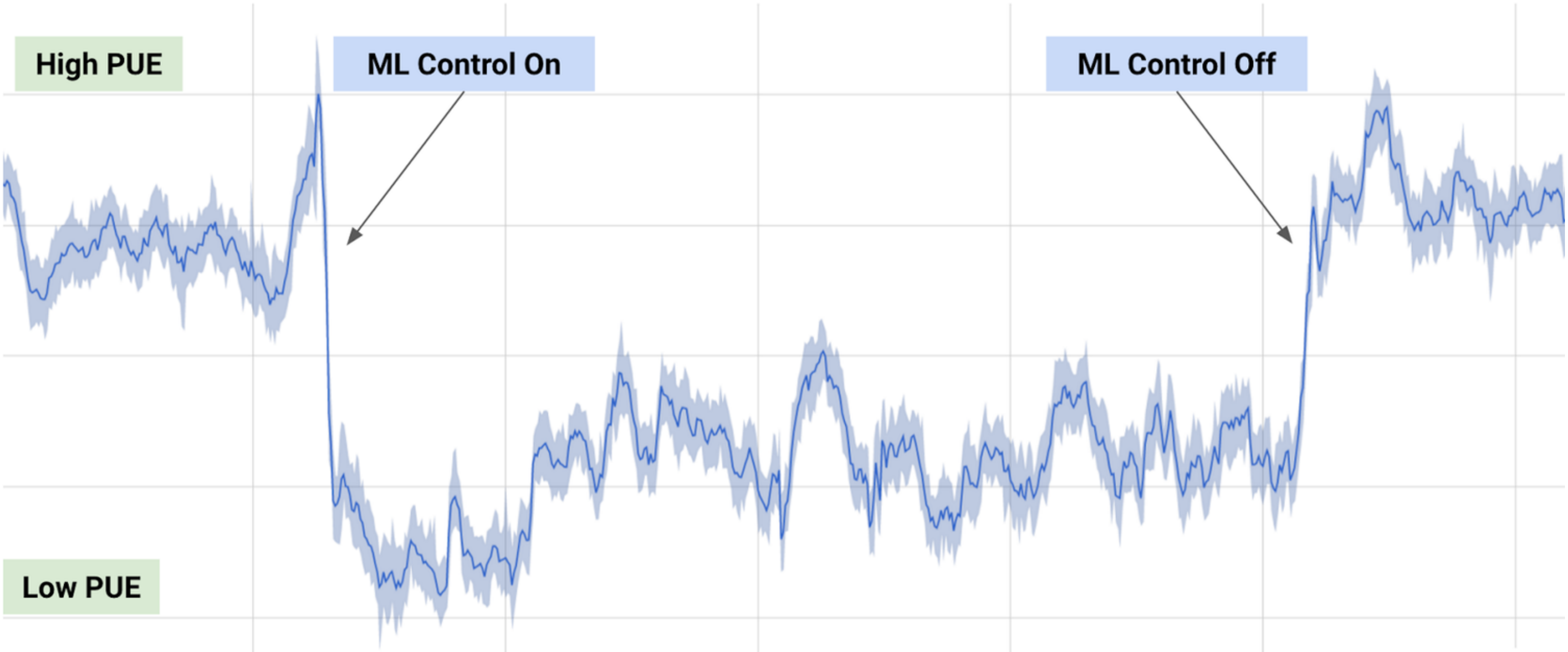
Autonomous (self-driving) computing systems

- Complex computing systems should have autonomic properties
- Independently take care of the regular maintenance and optimization tasks
- Reduce workload on the system administrators
- Distilled four properties of a self-managing (i.e. autonomic) system:
  - Self-configuration: adapting to dynamically changing environments
  - Self-optimization: tuning resources and balancing workloads to maximize use of IT resources
  - Self-healing: discovering, diagnosing, and acting to prevent disruptions
  - Self-protecting: anticipating, detecting, identifying, and protecting against attacks

Monitor-Analyse-Plan-Execute

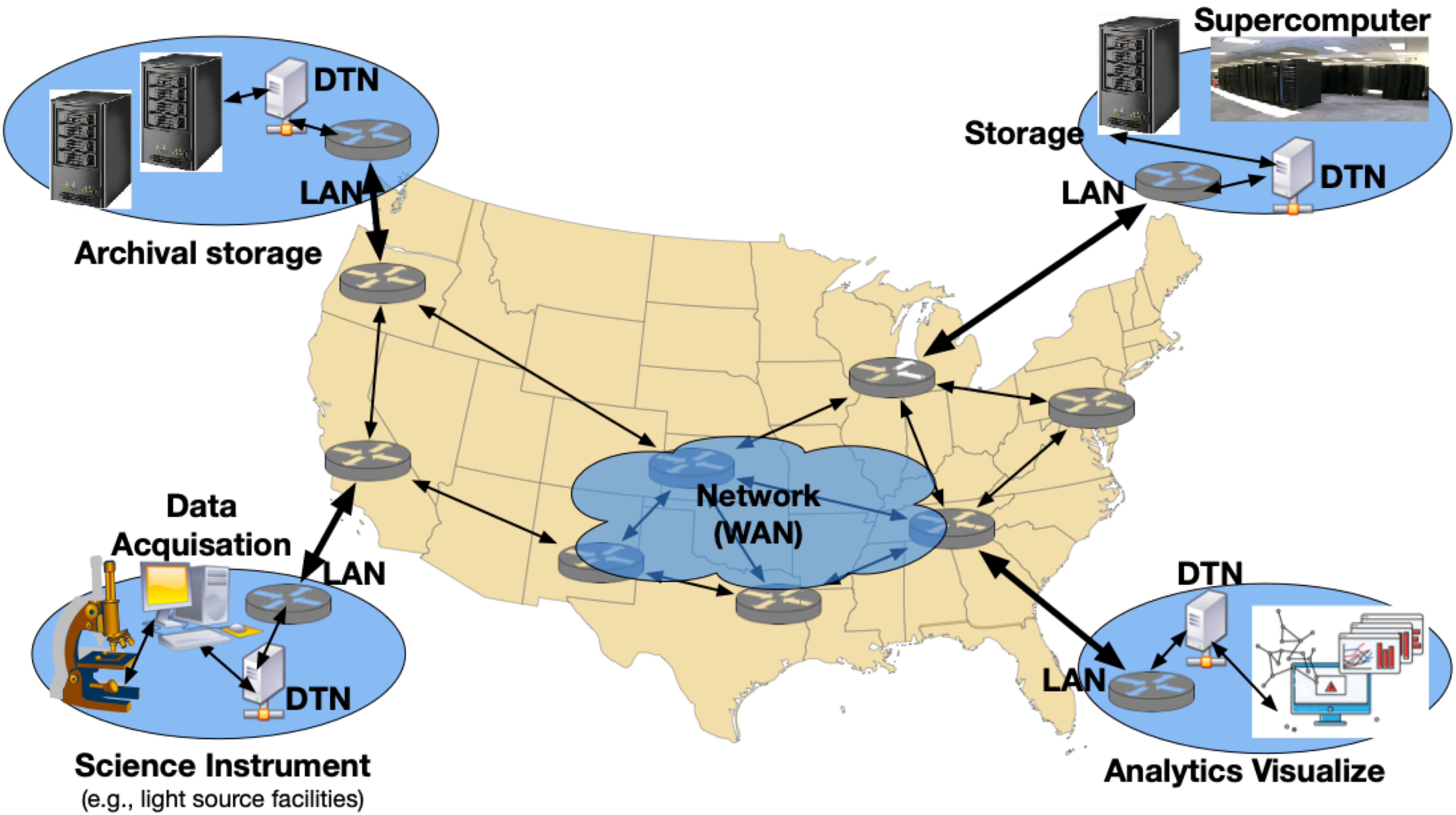


DeepMind AI Reduces Google Data Centre Cooling Bill by 40%



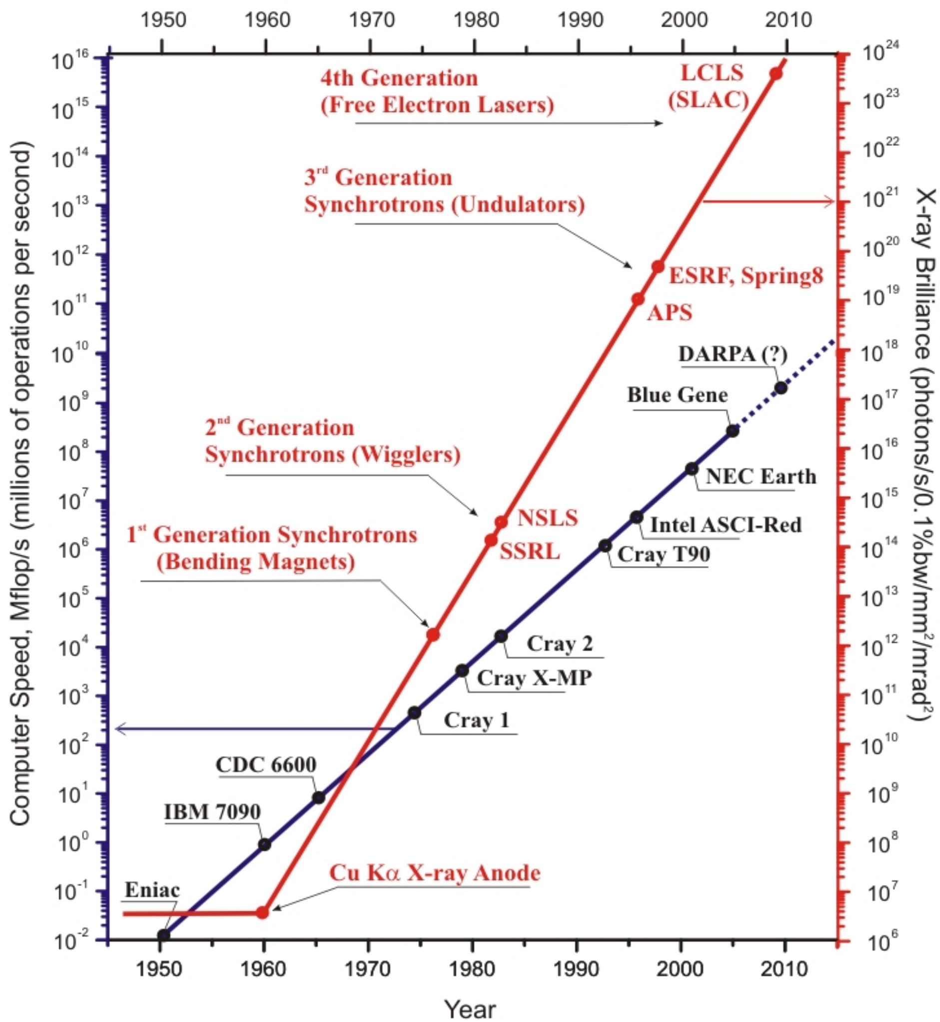
https://deepmind.com/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-40/

Distributed Science Ecosystem



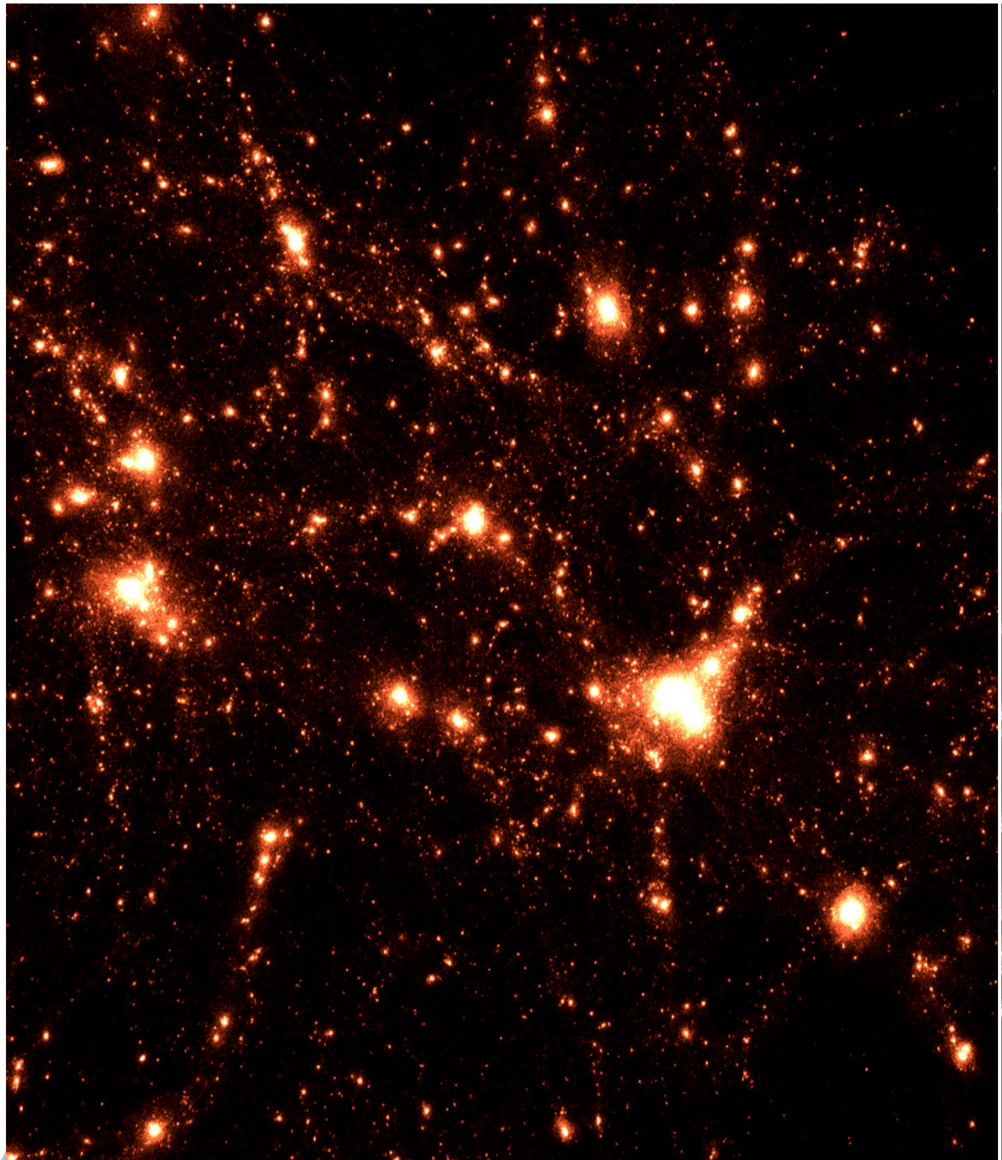
X-ray sources produce a lot of photons, which translates to a lot of data

Computer speed:
12 orders
of magnitude
in 6 decades

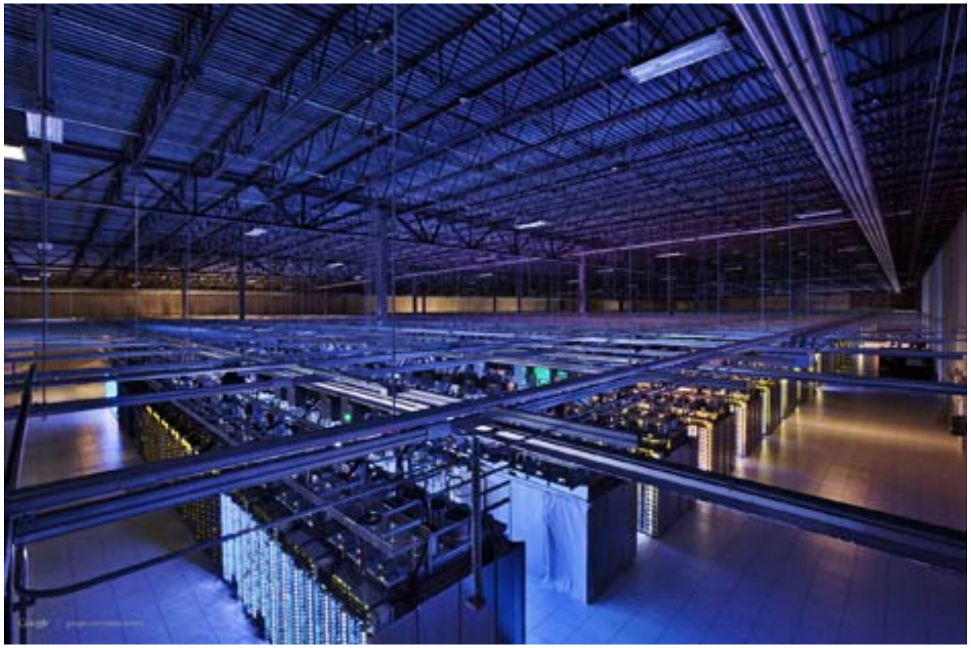


X-ray source
brilliance:
18 orders
of magnitude
in 5 decades!

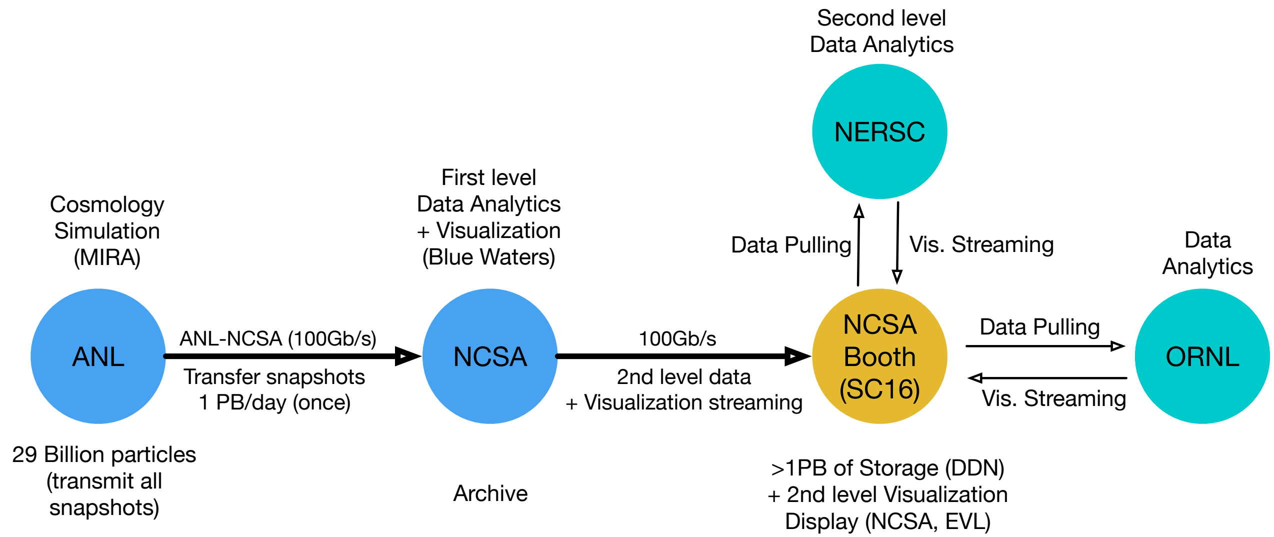
Huge amount of data from extreme scale simulations and experiments

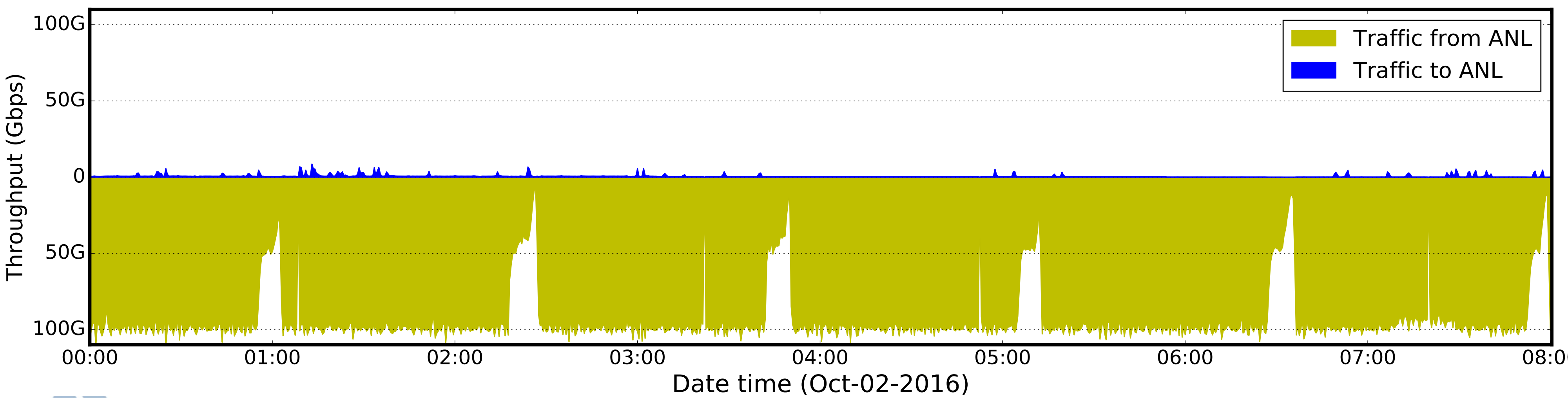



Systems have different capabilities

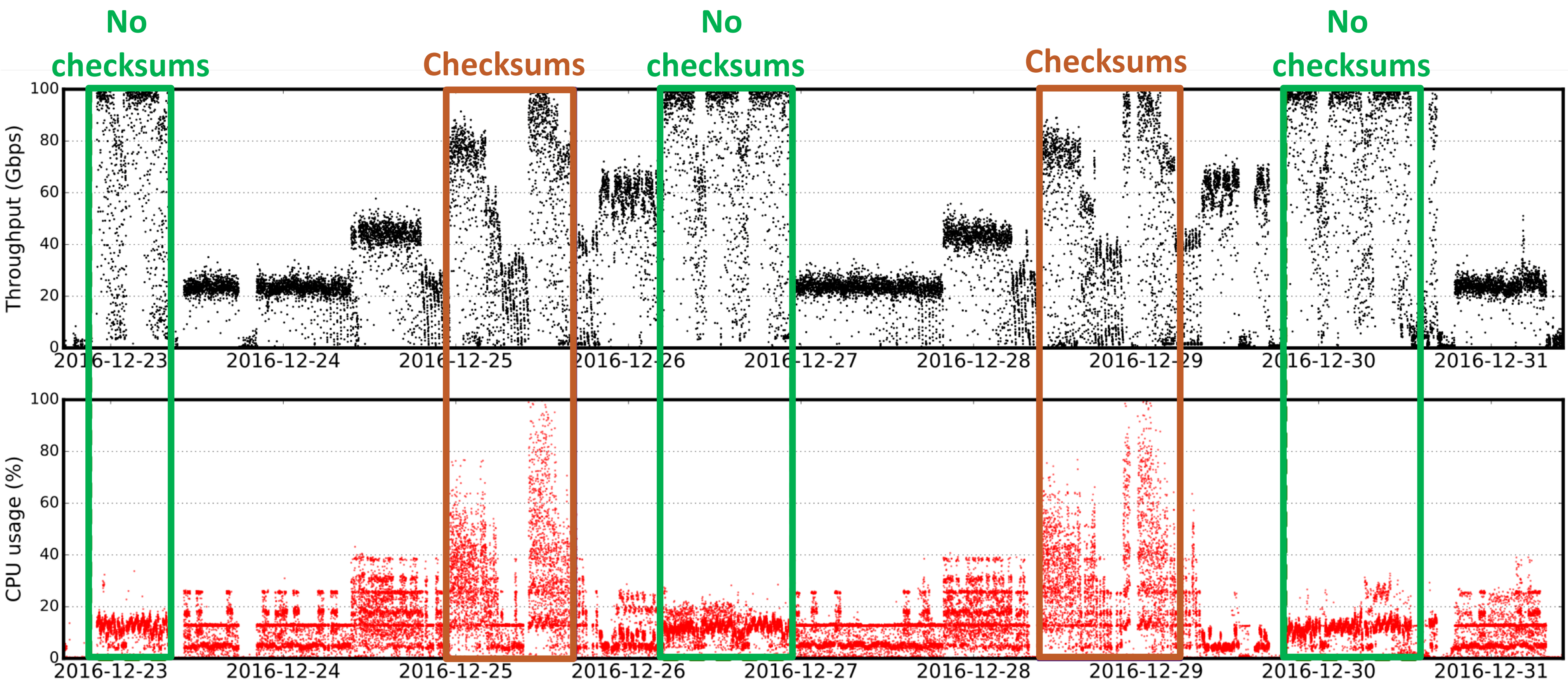


Transferring 1 PB in a day



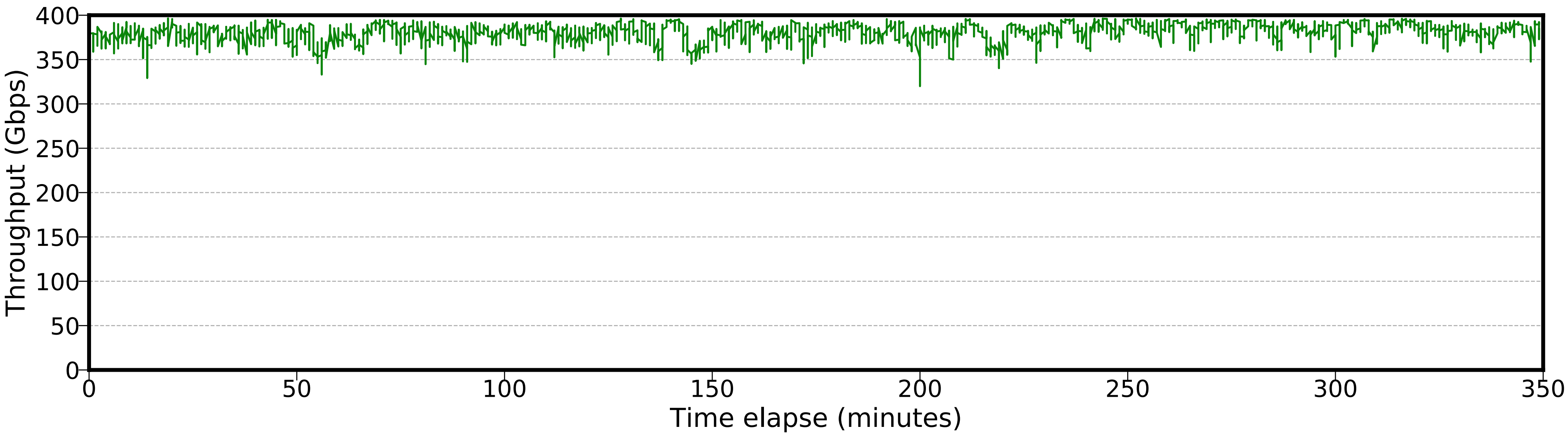


Identified infrastructure bottlenecks

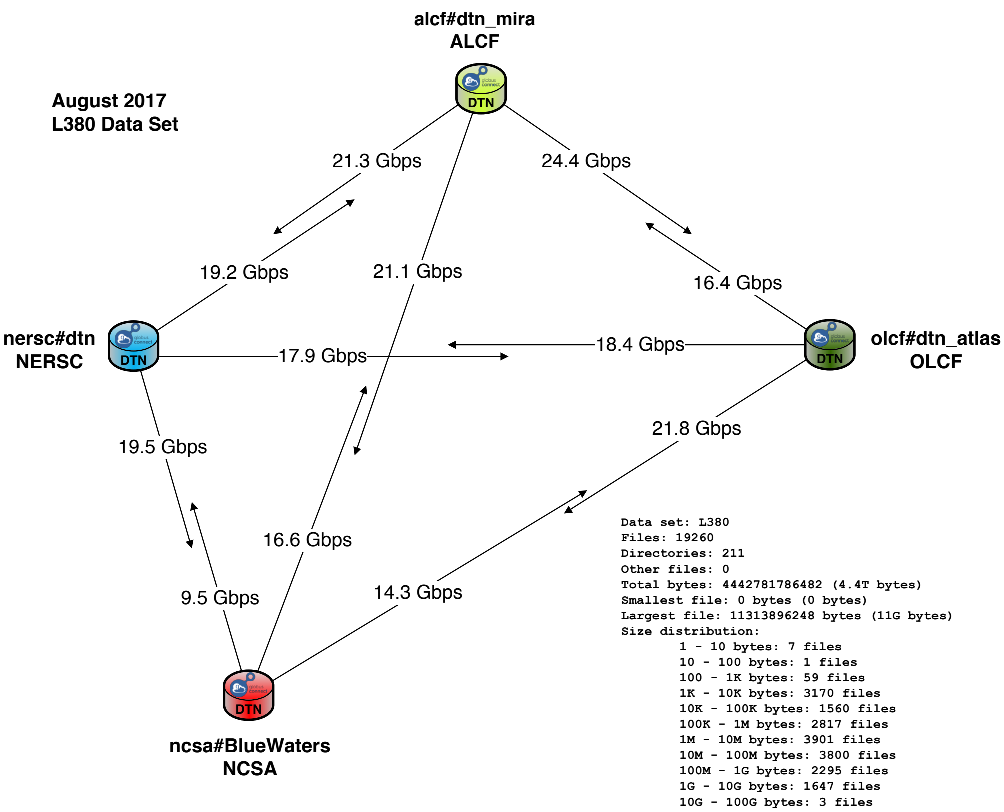


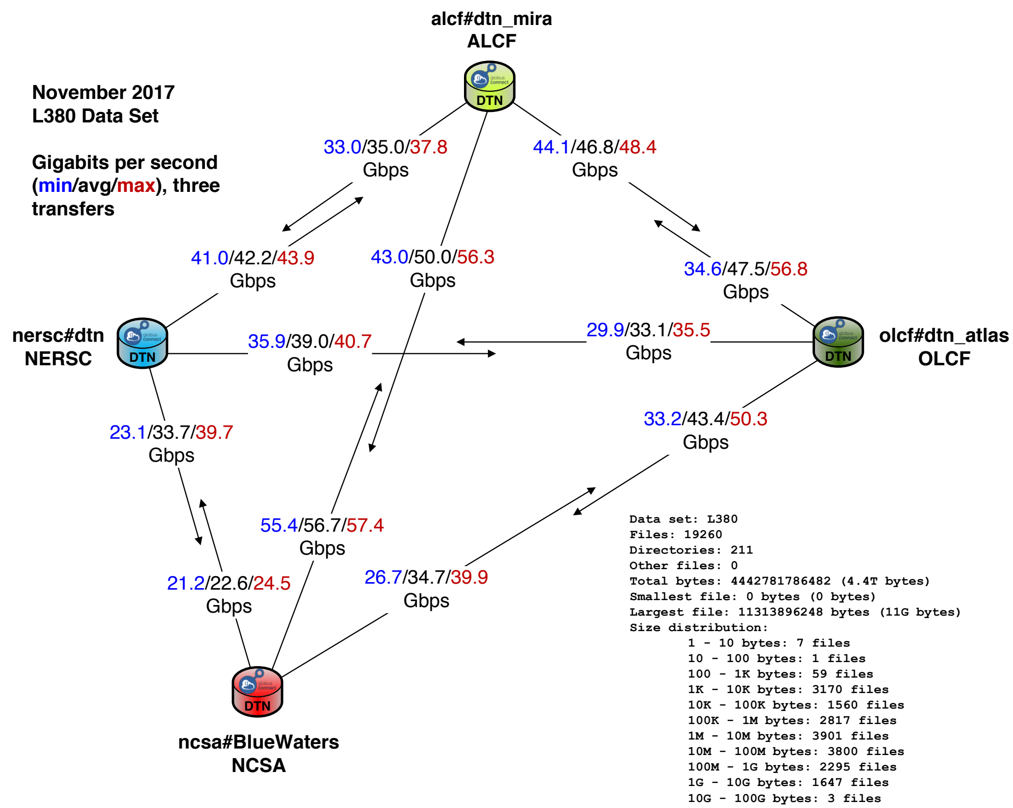
Outcome: ALCF upgraded data transfer nodes

Transferring 1 PB in 6 hours



Improved data transfer rates between supercomputer facilities by 1.5x to 3.5x

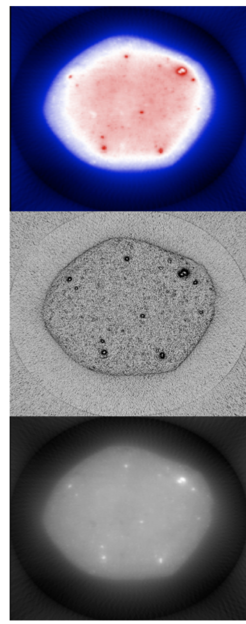




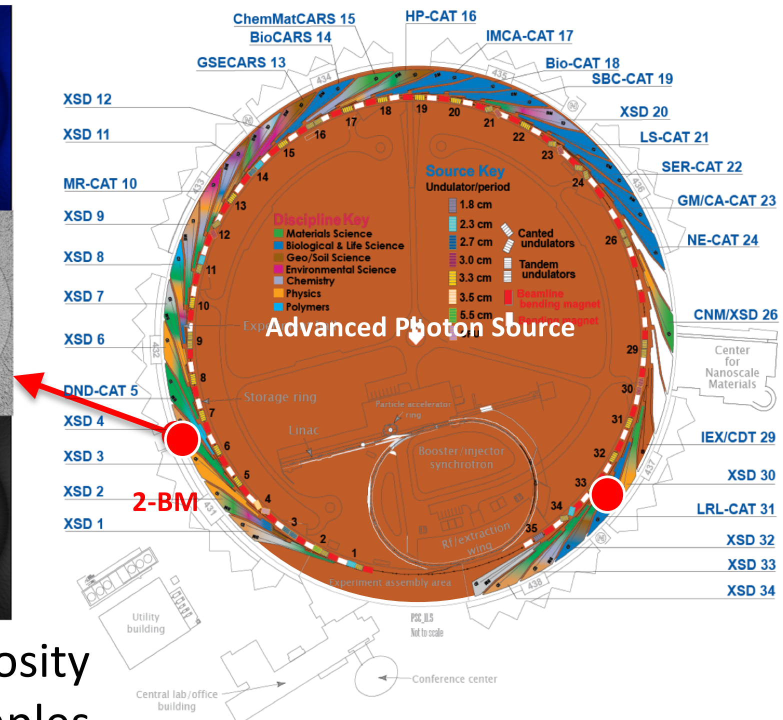
Scaled light source codes to leadership computers

X-ray nano/microtomography

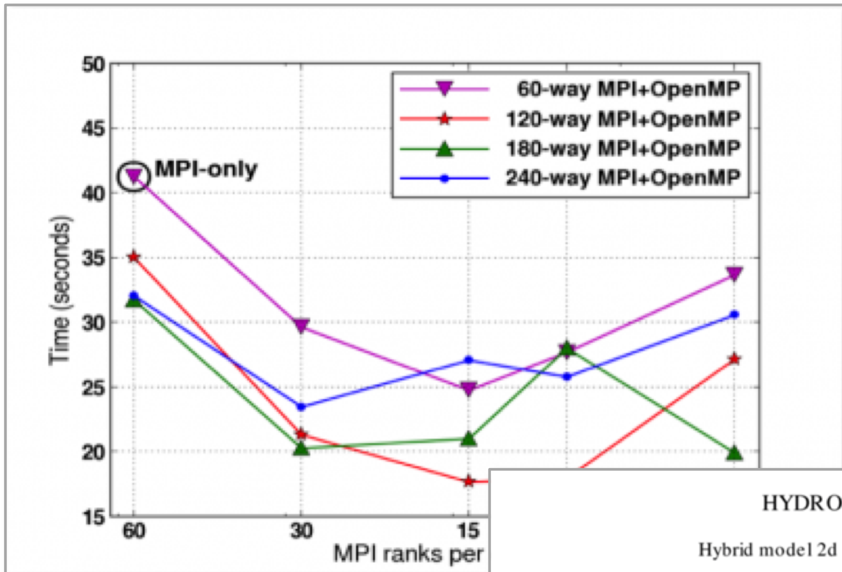
Bio, geo, and material science imaging. (Bicer, Gursoy, Kettimuthu, De Carlo, et al.). Innovative in-slice parallelization method permits reconstruction of 360x2048x1024 dataset in ~1 minute, using **32K BG/Q cores**, vs. many days on typical cluster: enables quasi-instant response

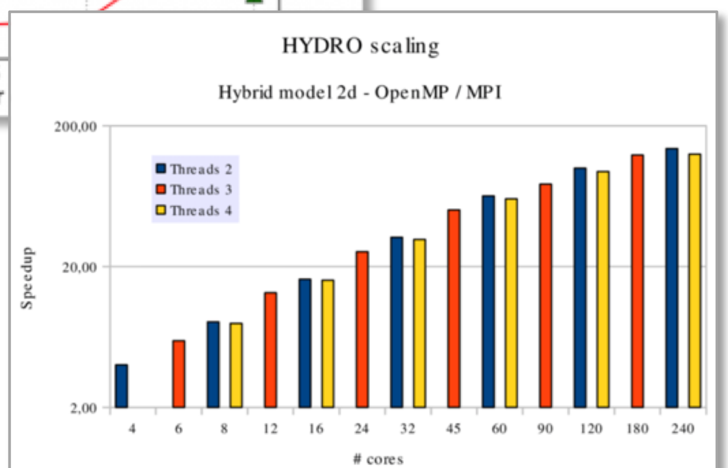


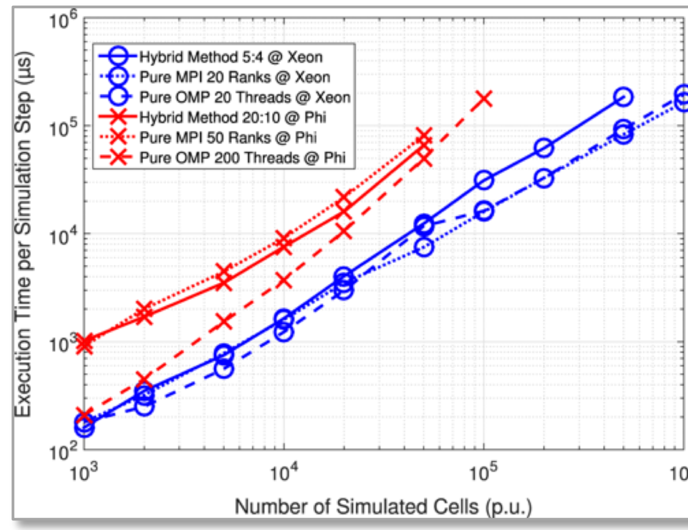
Micrometer porosity structure of shale samples



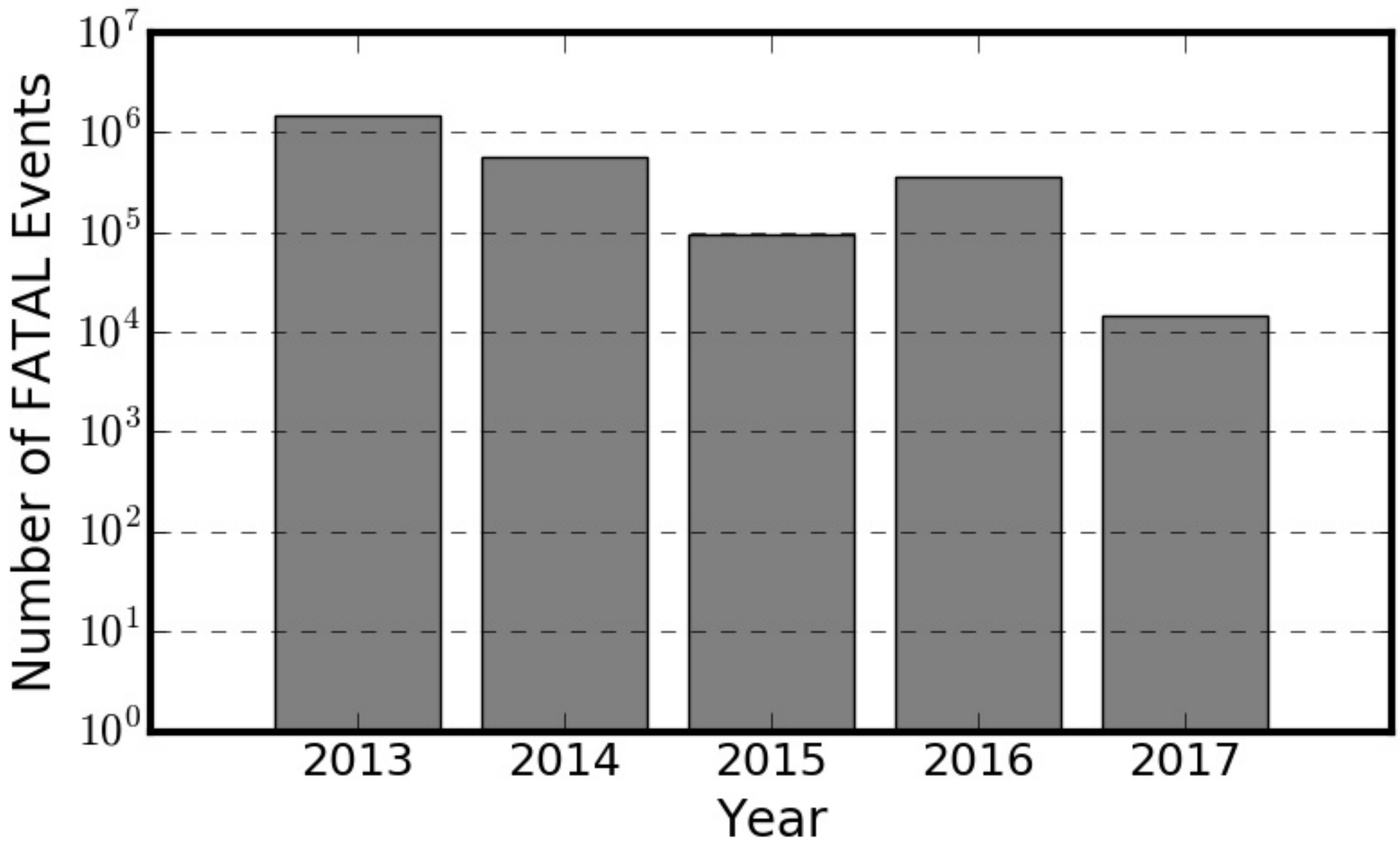
Why are we still tuning?



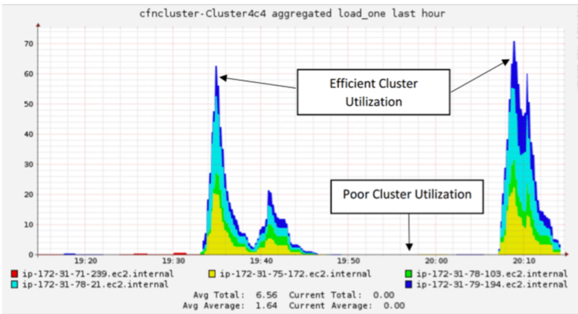


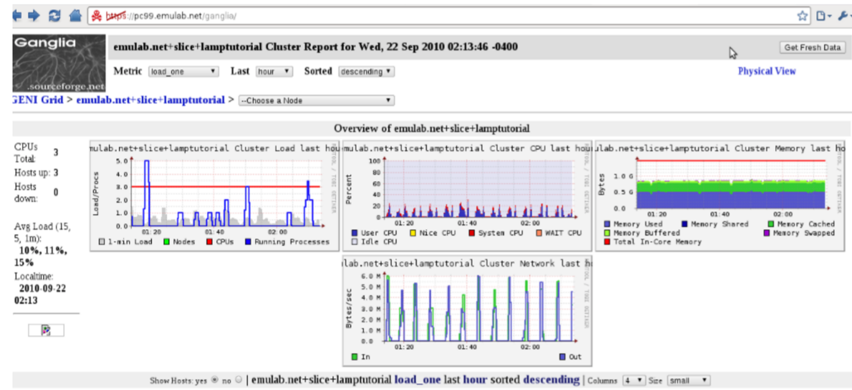


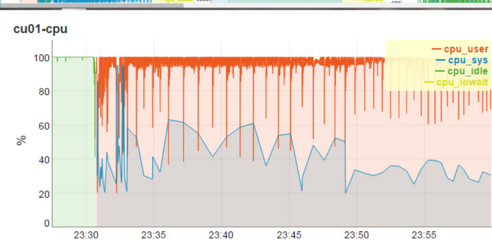
Core hours lost due to suboptimal failure management



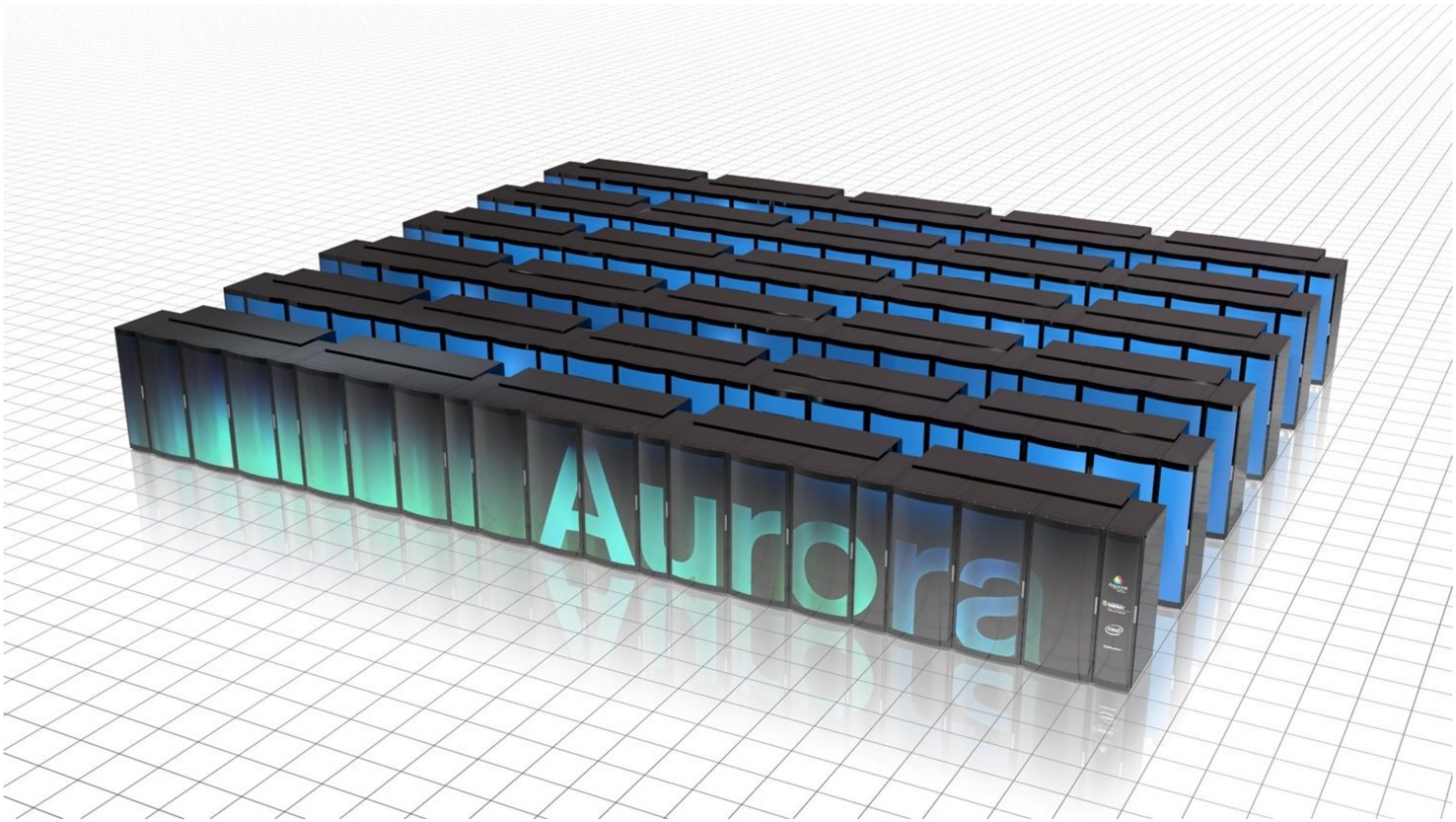
Why are we still troubleshooting?







Why are we still designing?

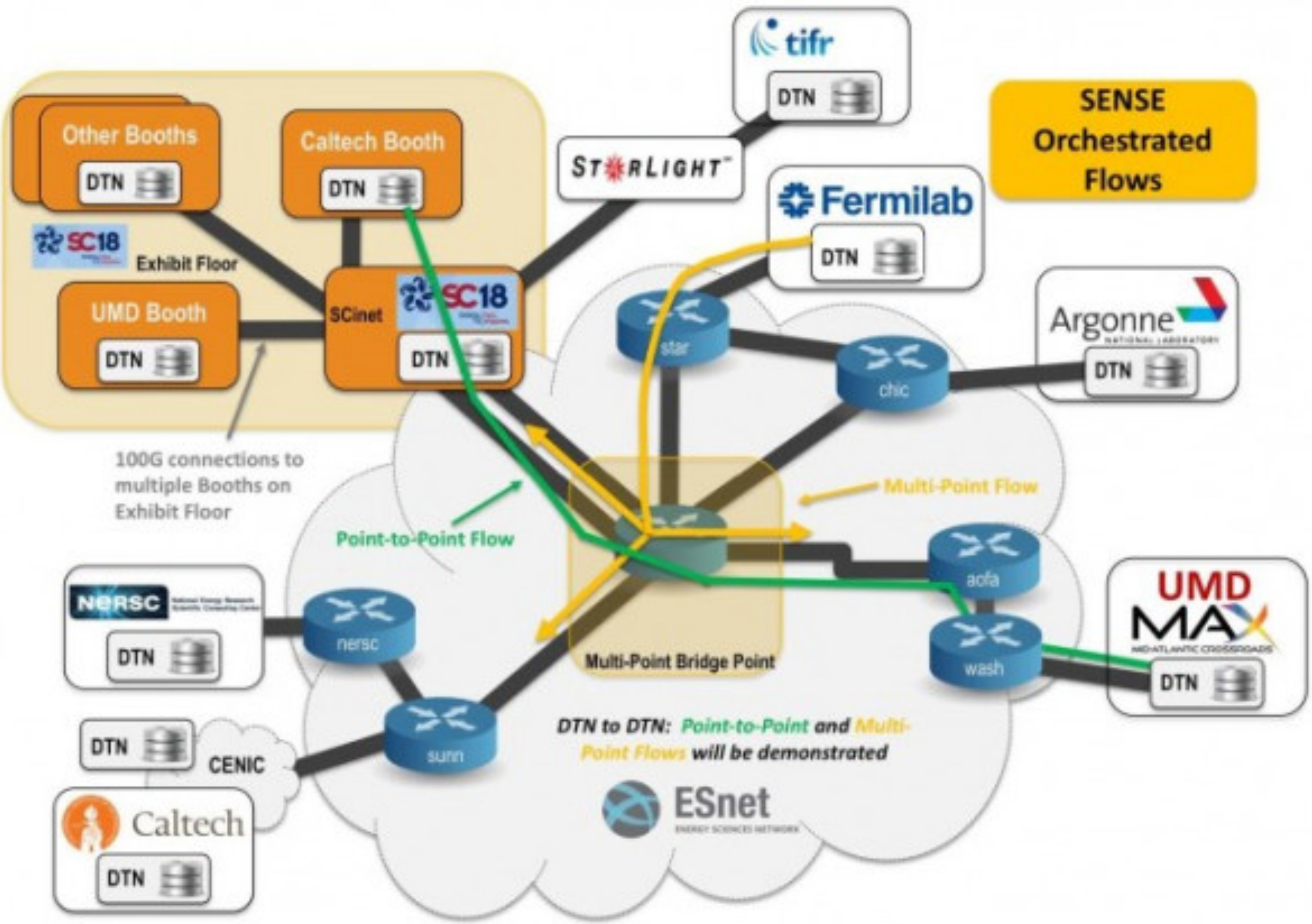


1920 telephony

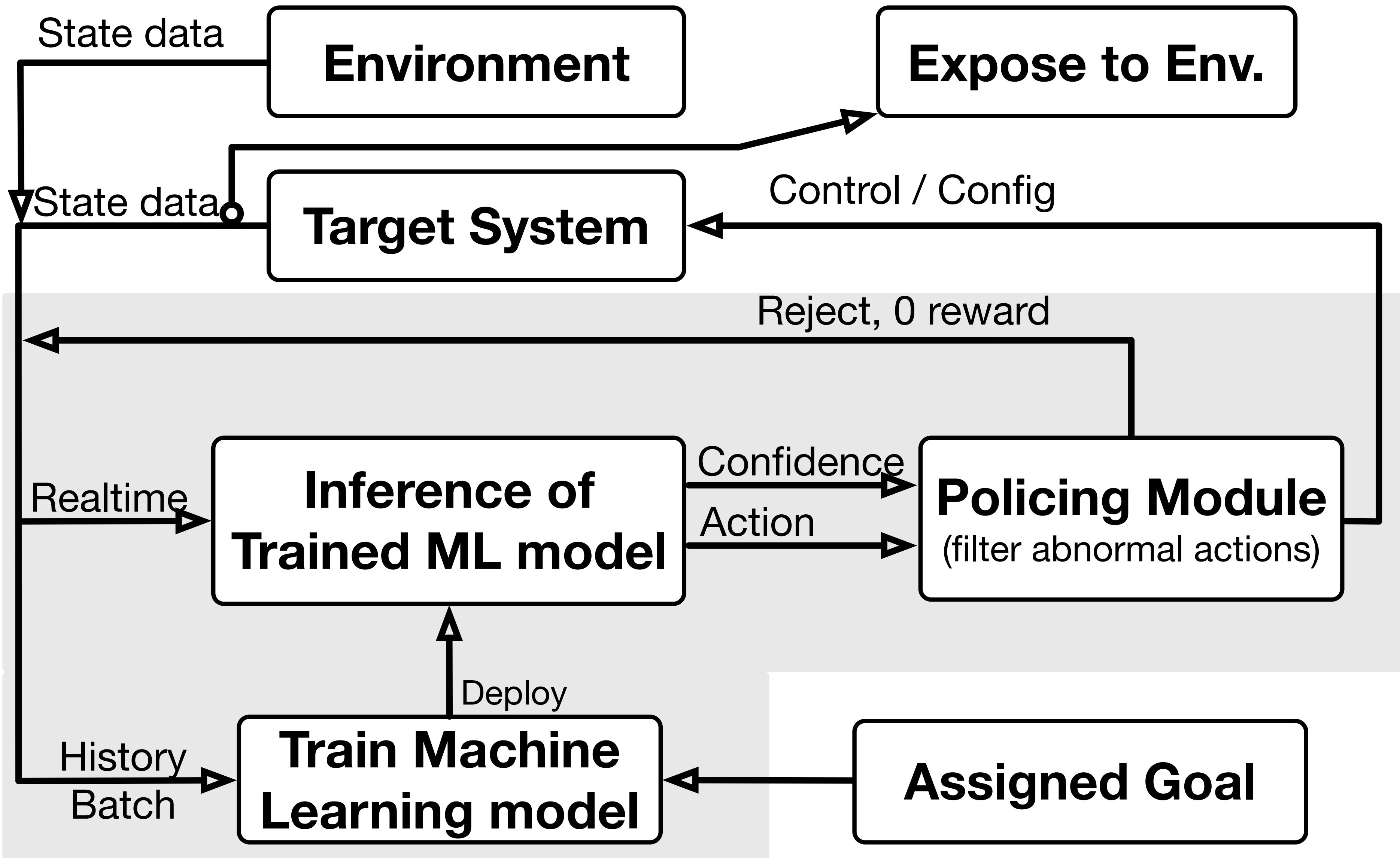
Autonomous (smart) system

- Definition: achieves a given goal by tuning the parameters appropriately based on environmental conditions without human intervention
  - Ability of self awareness - learns the relationship between a performance metric, tuning parameters and the environmental conditions
  - Strive to achieve a goal for that performance metric
- How does it differ from automation?
  - Automation is an execution of predefined rules
  - Automation works in a given environment but has to be adapted if the environment changes
  - Autonomous system adapts on its own to a changing environment

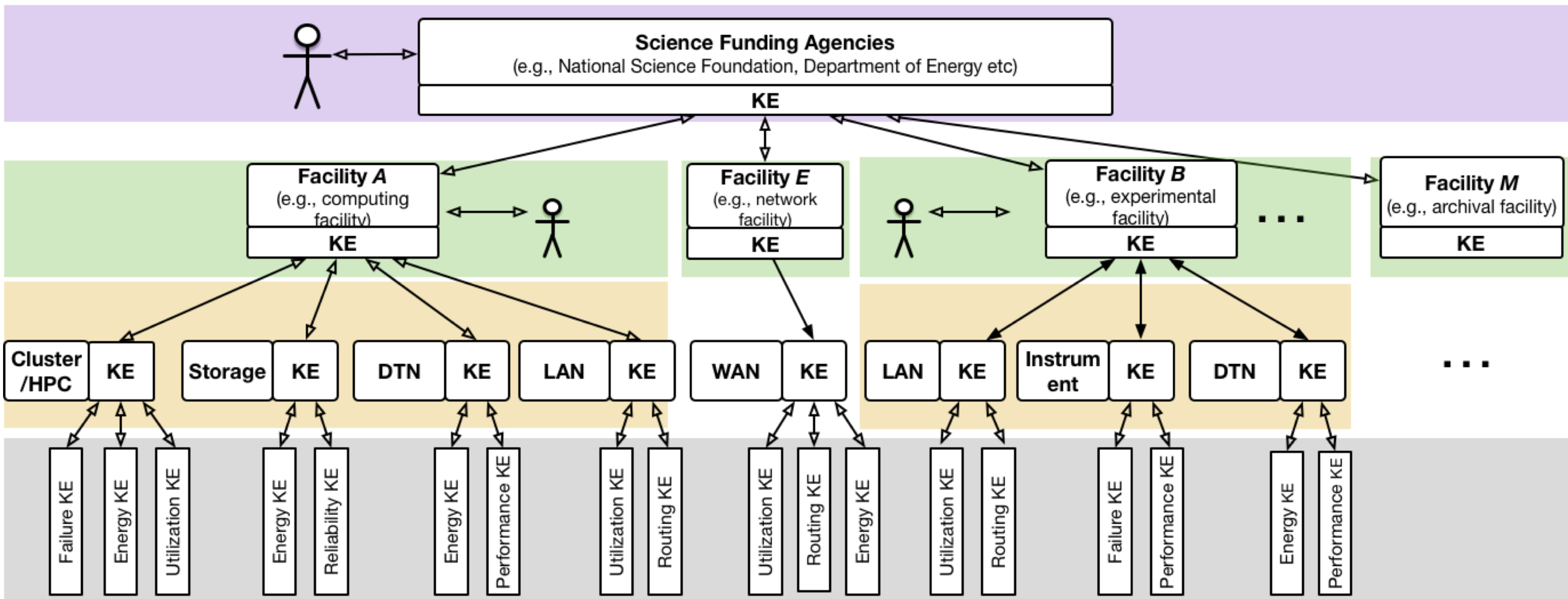
Autonomous network



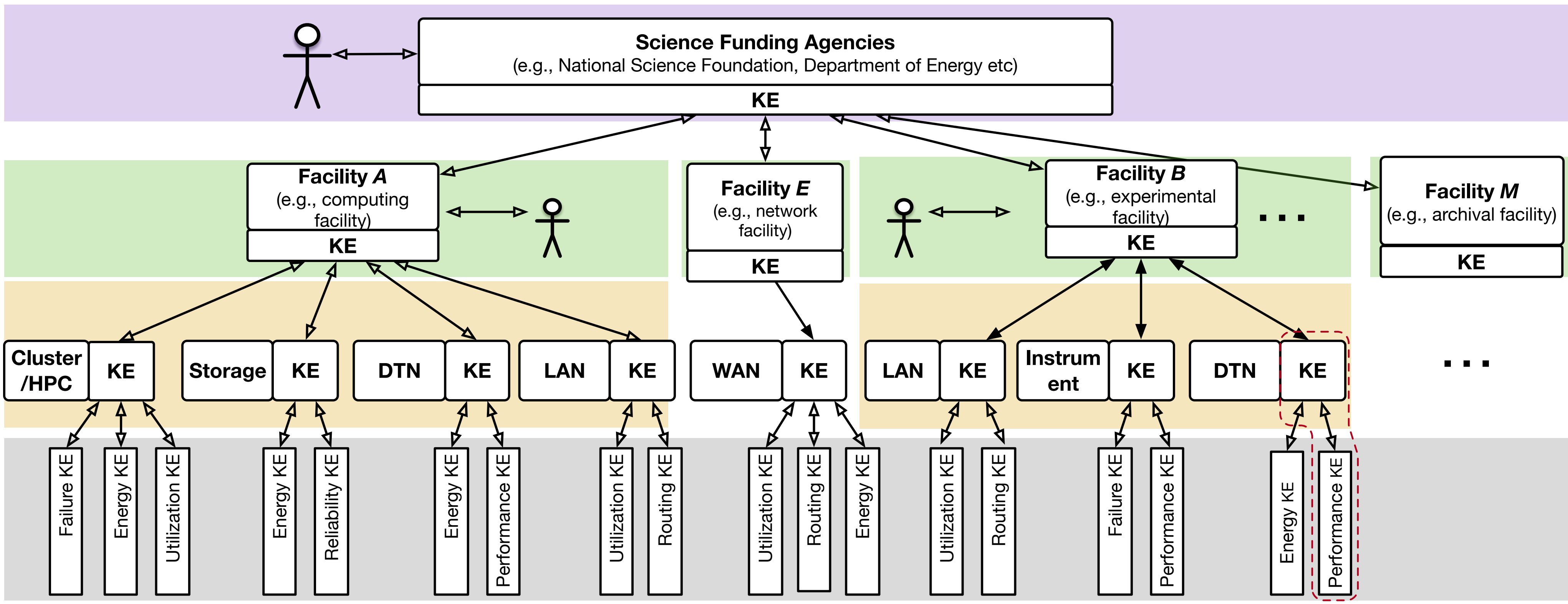
Architecture of an individual autonomous system



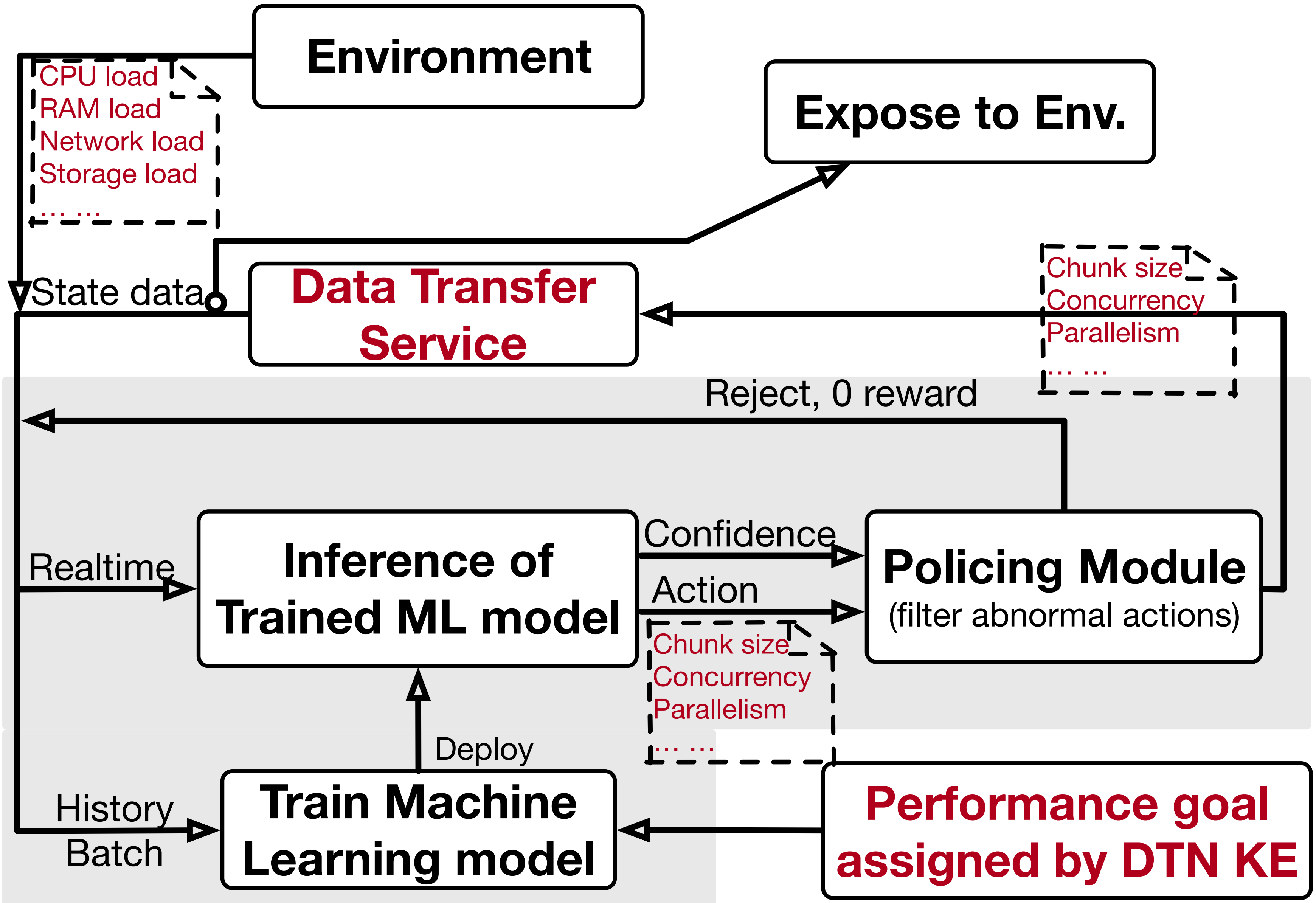
Autonomous science infrastructure architecture



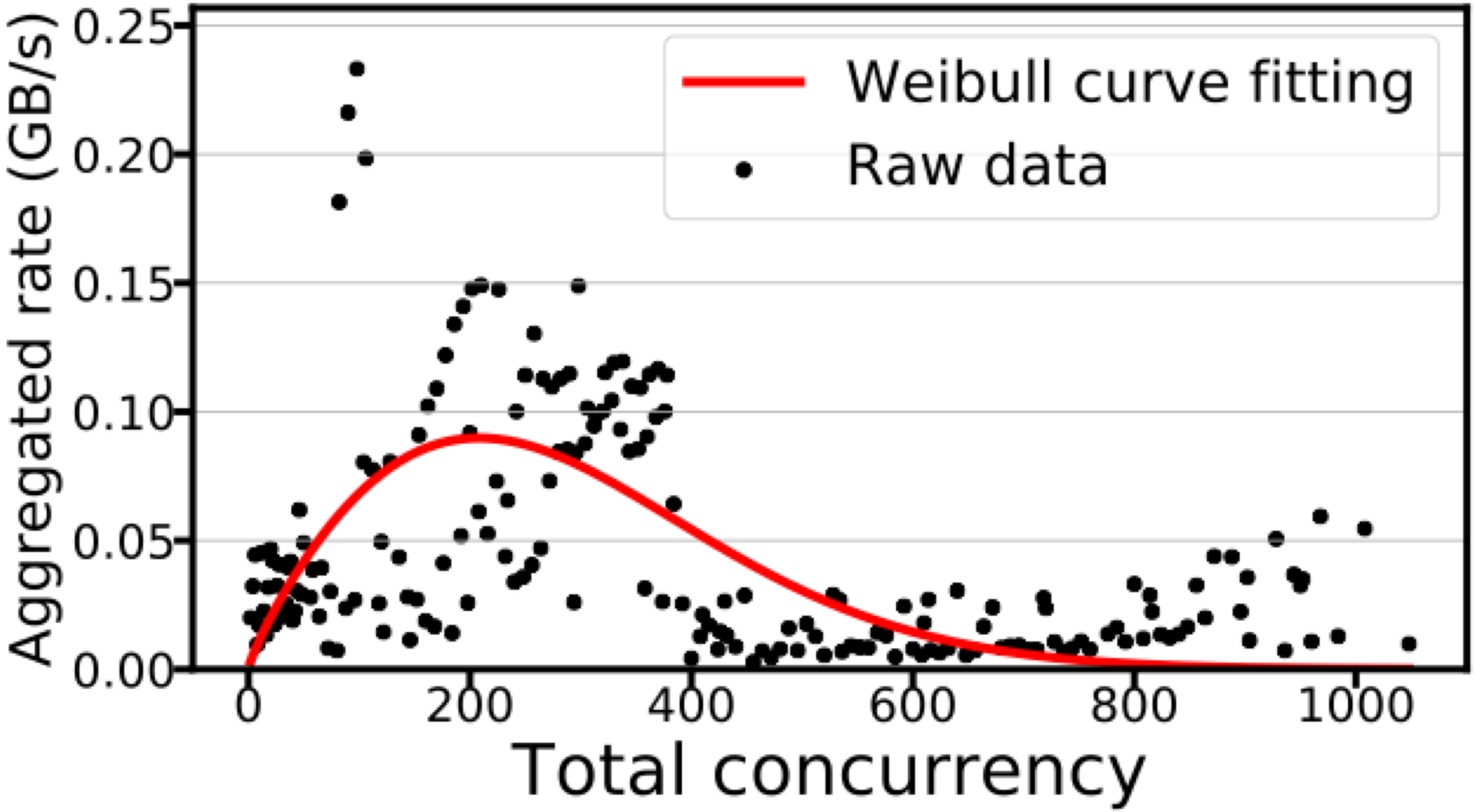
Autonomous science infrastructure architecture



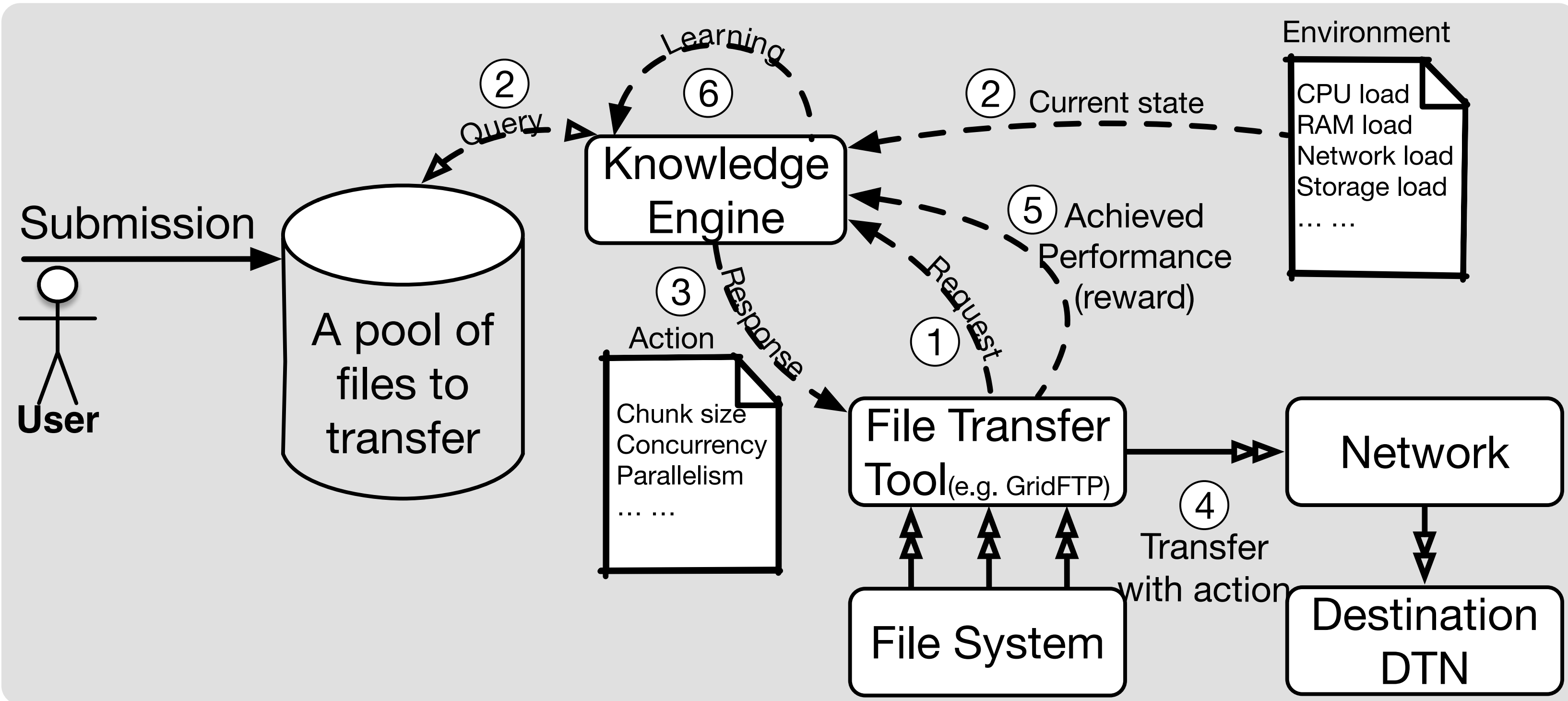
Autonomous performance module for DTN



Concurrent transfers

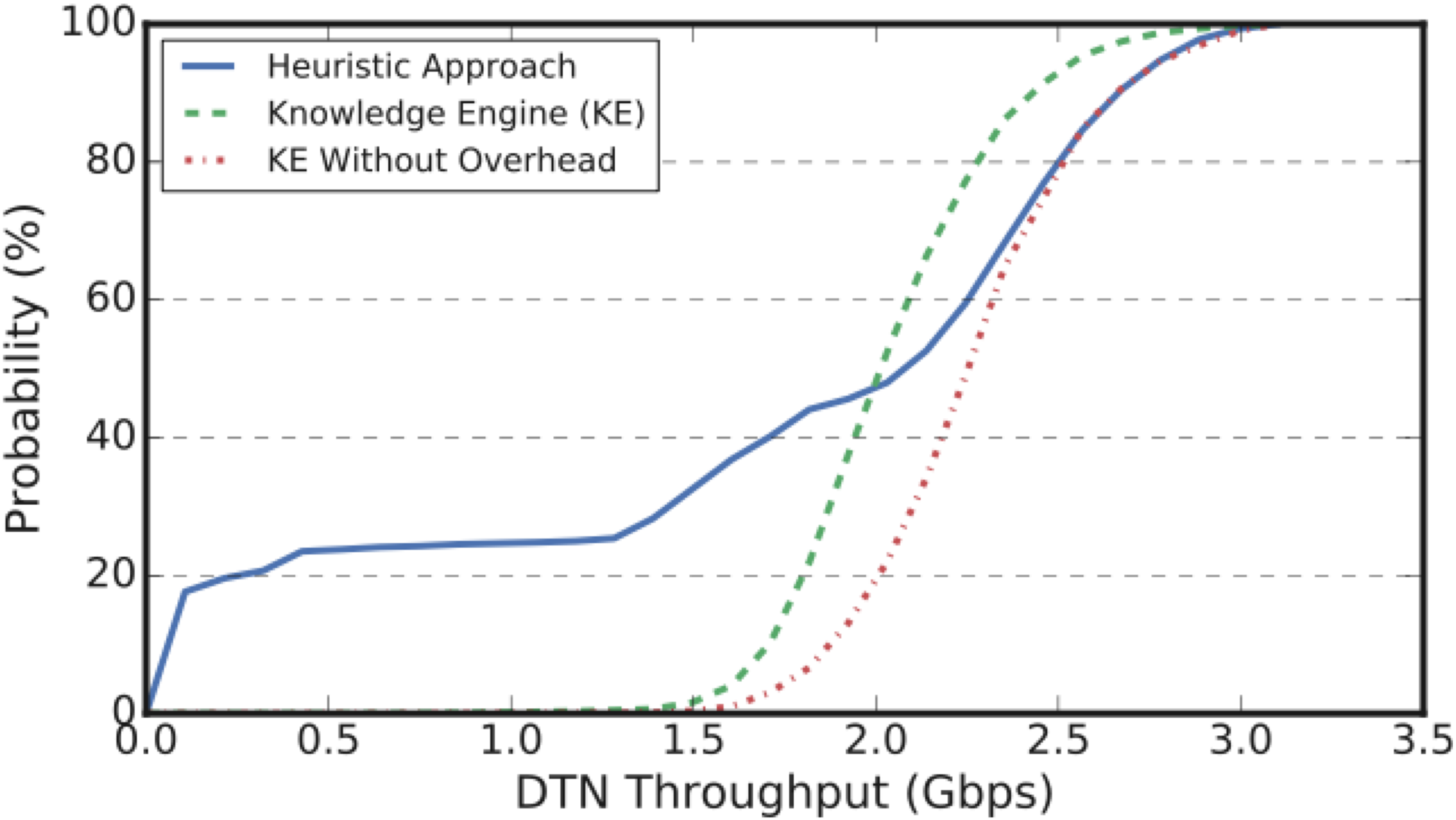


Smart data transfer node



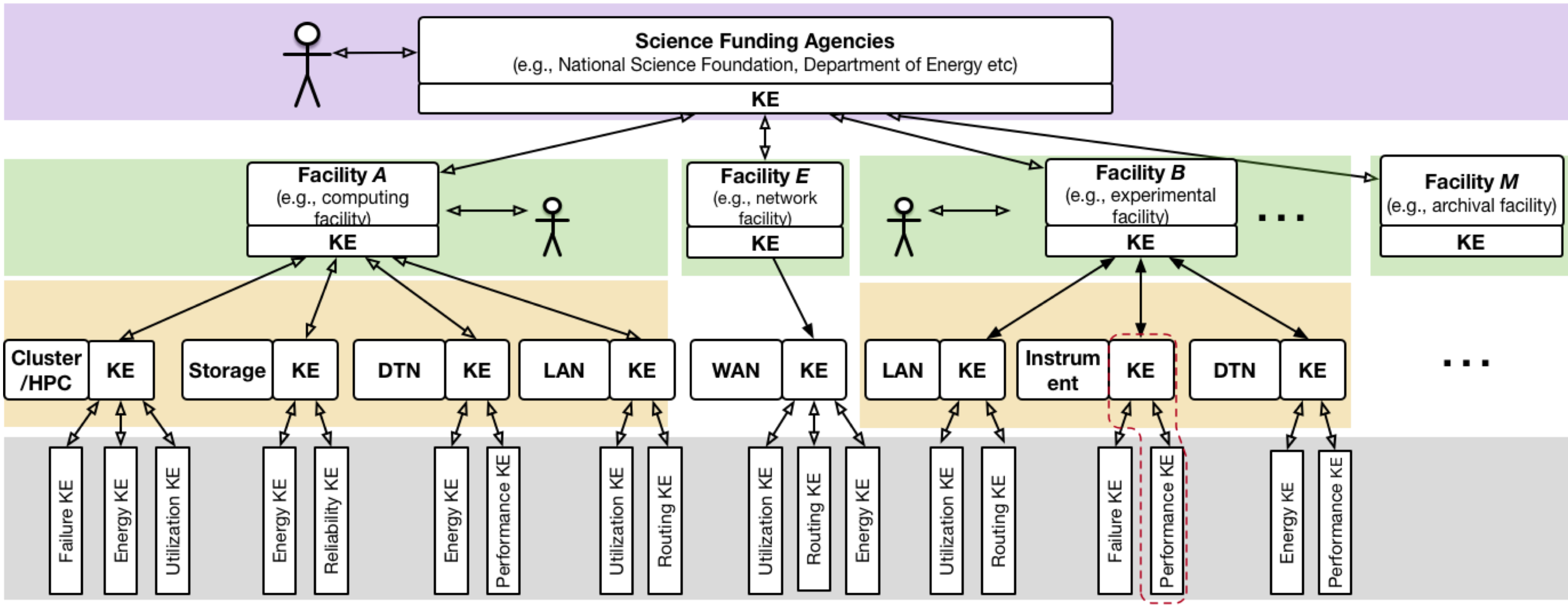
① A file transfer tool requests a file to transfer from the KE. The KE ② checks the current DTN state and ③ responds to the transfer tool with a chunk of file and corresponding optimal transfer parameters (the steering action). ④ The transfer tool transfers the associated chunk with the parameters and monitors the aggregate DTN throughout during this transfer. ⑤ Once completed, DTN's average aggregate throughput is reported to the KE as a reward for its actions. ⑥ Based on the reward (encourage or discourage), the KE updates its internal model parameters to improve its decision policy.

Results

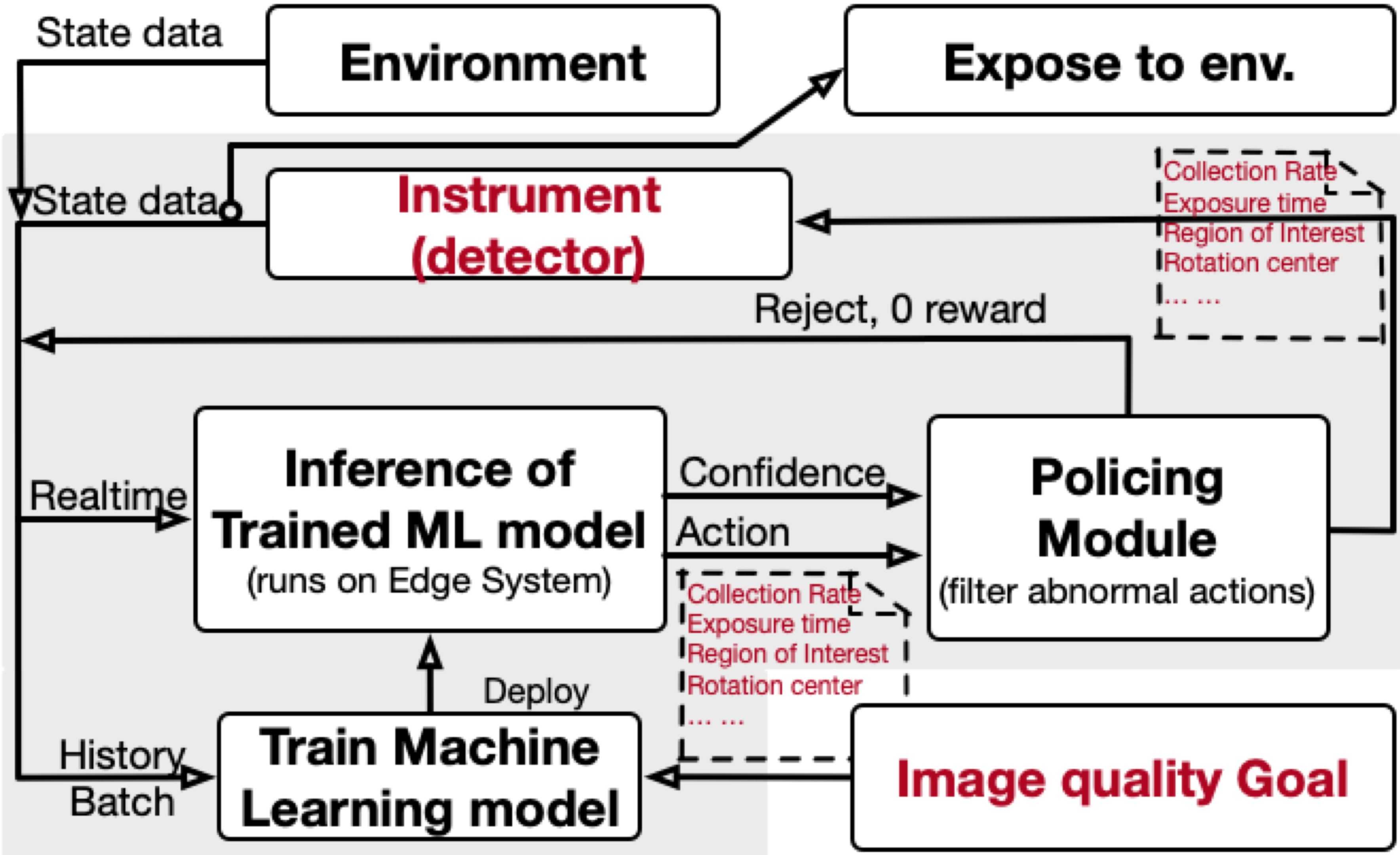


Heuristic Approach
Knowledge Engine (KE)
KE Without Overhead
Probability (%)
100
80
60
40
20
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
DTN Throughput (Gbps)

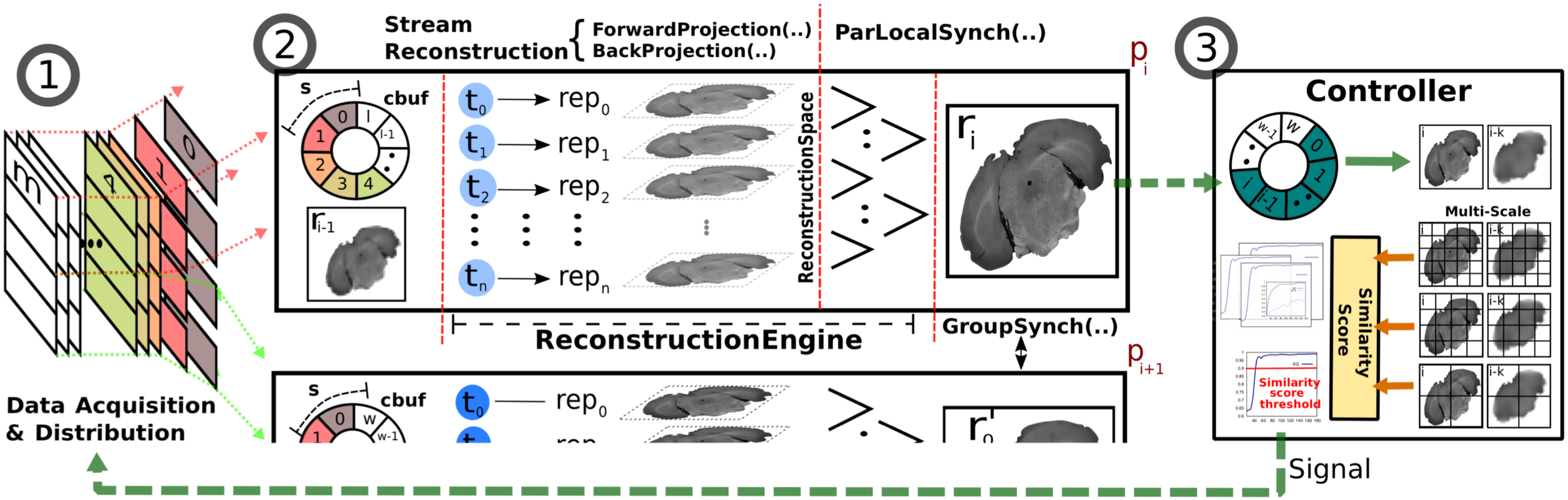
Autonomous science infrastructure architecture



Smart instrument – self configure and self optimize

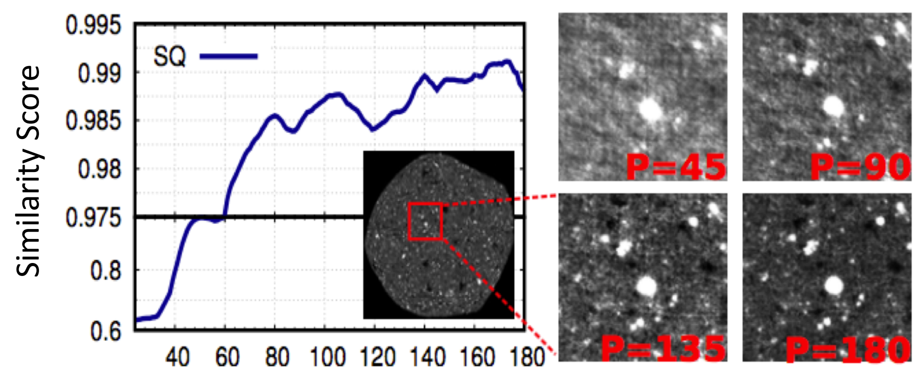


Experiment monitoring and steering



Autonomous stream processing system that allows data streamed from beamline computers to be processed in real time on a remote supercomputer with a control feedback loop used to make decisions during experimentation

– Reduce data acquisition time by 22–44% for the datasets considered in our experiments

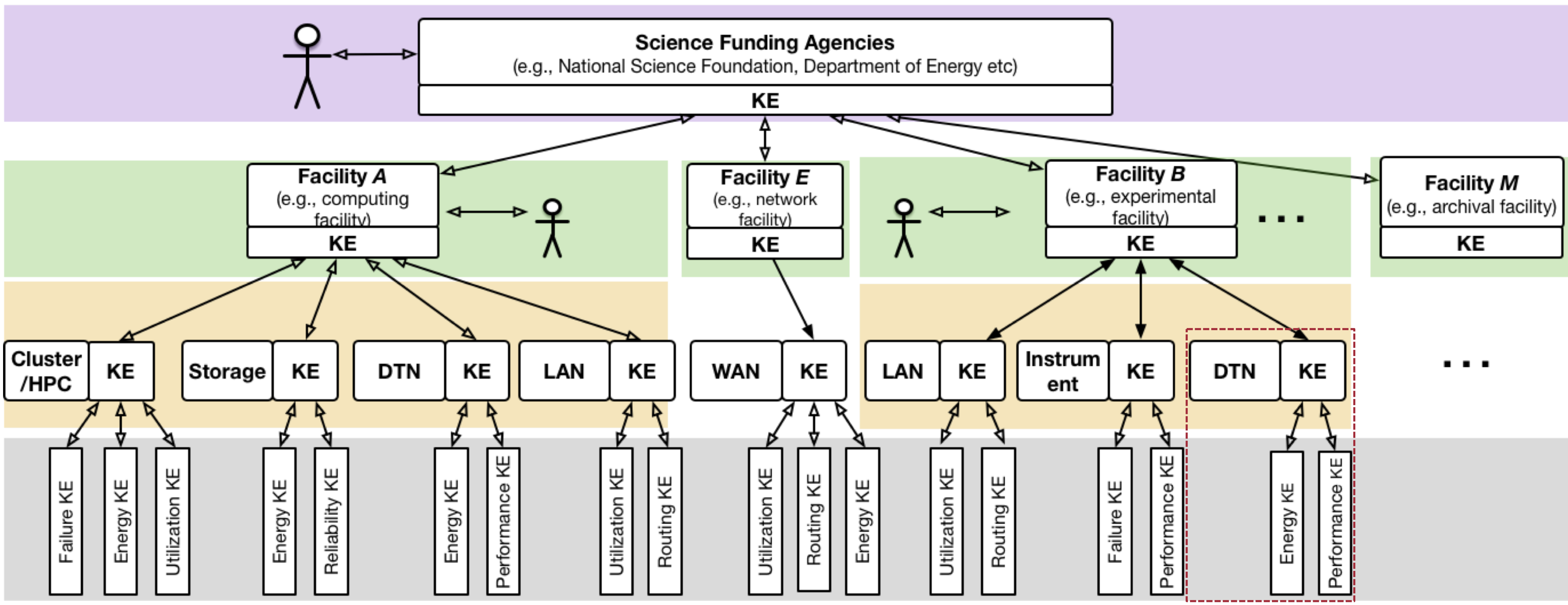


Streamed Projections

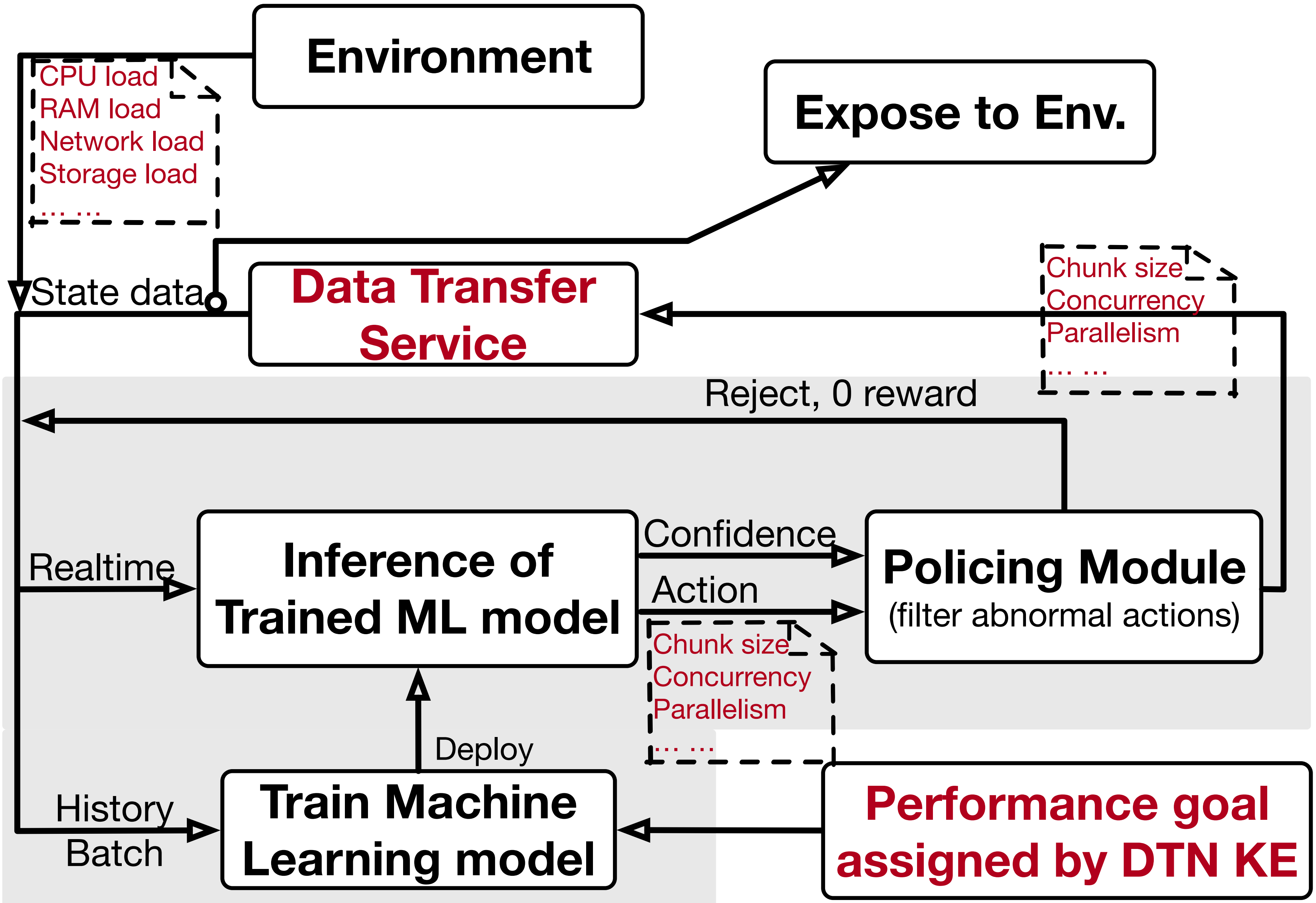
Reconstructed Image Sequence

Image Quality with respect to Streamed Projections

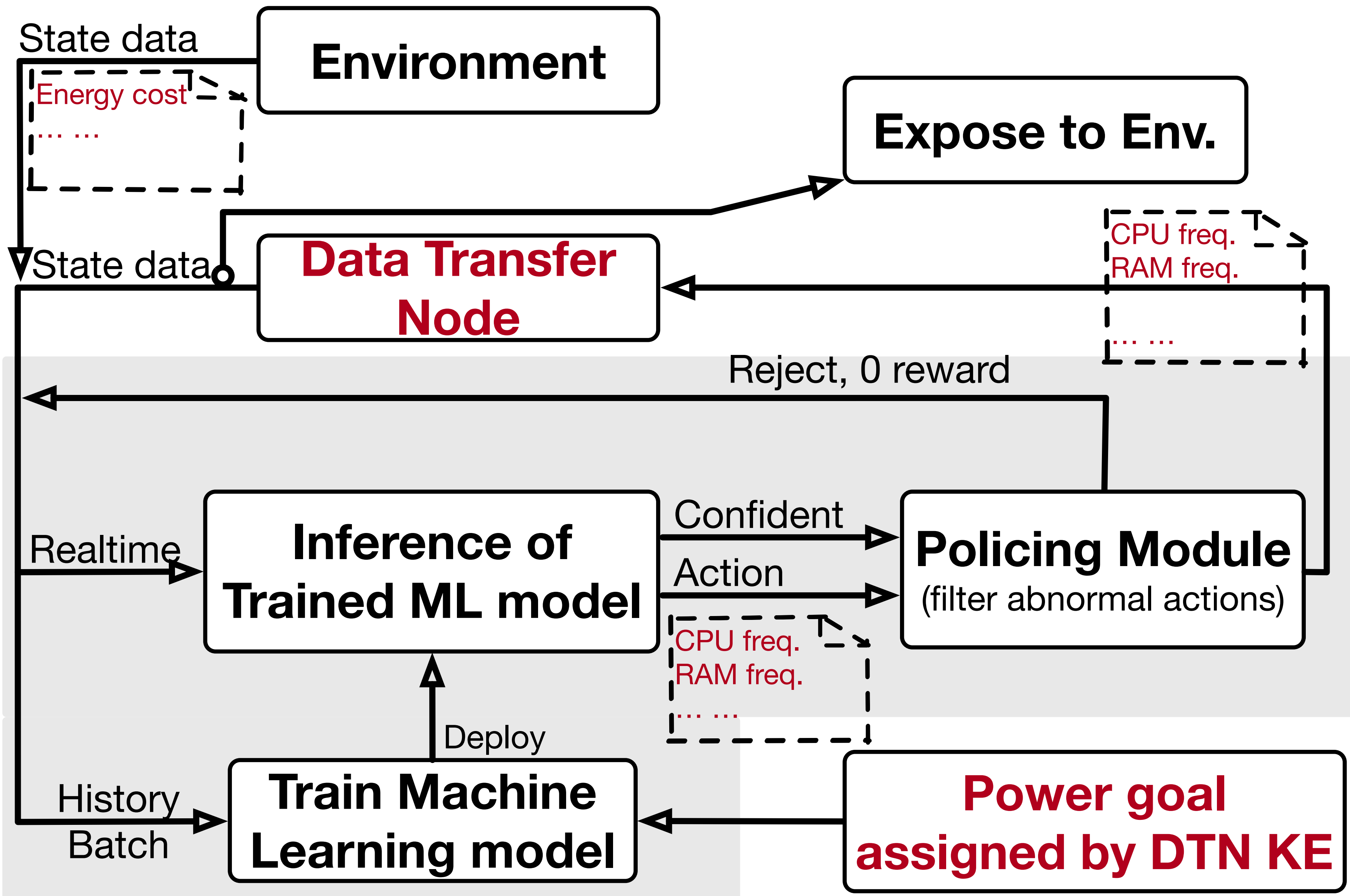
Autonomous science infrastructure architecture



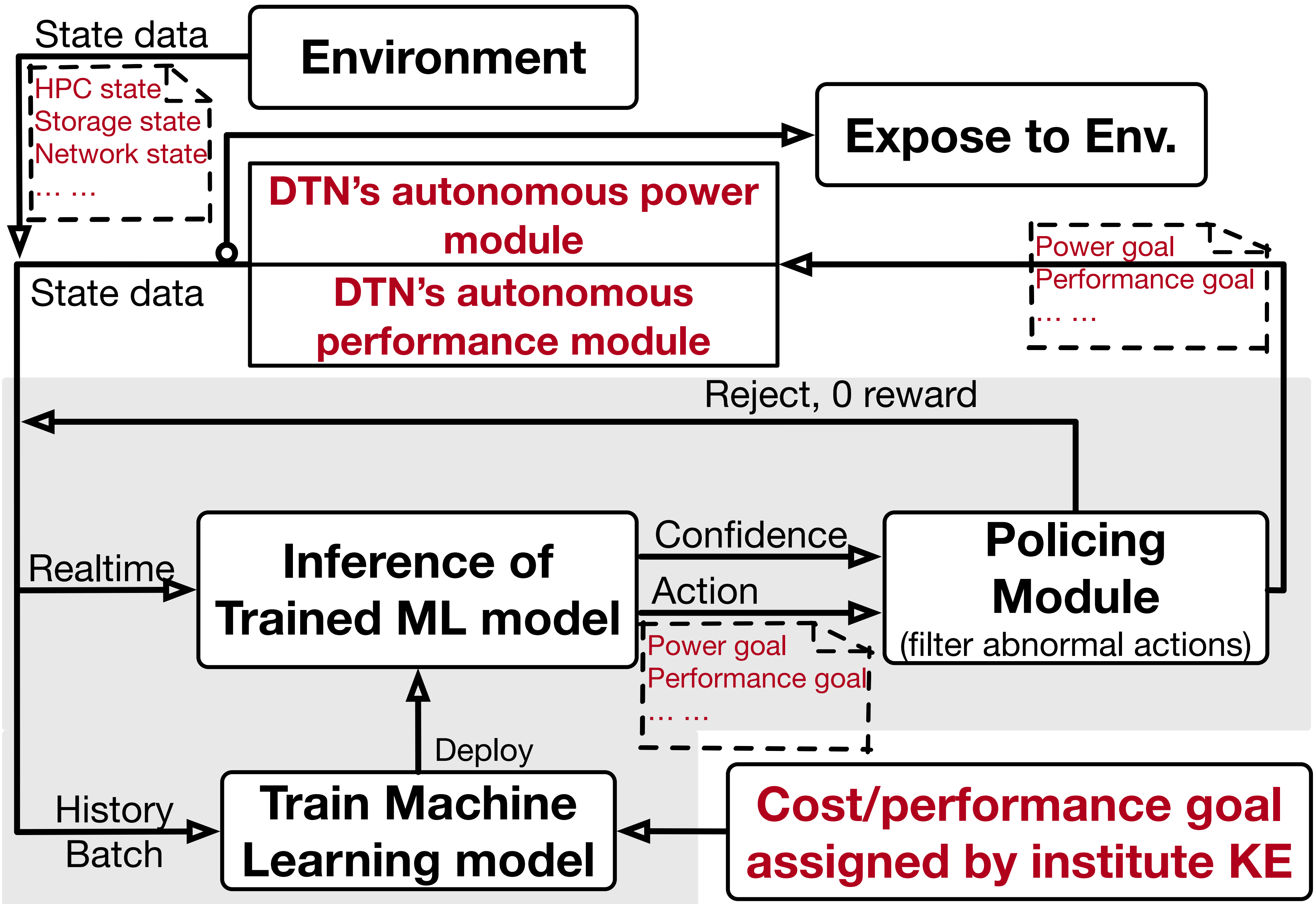
Autonomous performance module for DTN



Autonomous power module for DTN



Autonomous DTN



Acknowledgment

U.S. DEPARTMENT OF
ENERGY

Office of Science

Questions